WHAT IS CLAIMED IS:

1.   A method of adapting an n-gram language model for a new domain, the method comprising:

      receiving background data indicative of general text phrases not directed to the new domain;

      receiving a set of semantic entities used in the new domain and organized in classes;

      generating background n-gram class count data based on the background data and the semantic entities and classes thereof; and

      training a language model based on the background n-gram class count data.

2.   The method of claim 1 and further comprising:

      receiving adaptation data indicative of text phrases used in the new domain;

      generating adaptation n-gram class count data based on the adaptation data and the semantic entities and classes thereof; and

      wherein training the language model comprises training based on the background n-gram class count data and the adaptation n-gram class count data.

3.   The method of claim 2 and further comprising:

> generating background n-gram word data based on the background n-gram class count data and the semantic entities and classes thereof;
>
> generating adaptation n-gram word data based on the adaptation n-gram class count data and the semantic entities and classes thereof; and
>
> wherein training the language model based on the background n-gram class count data and the adaptation n-gram class count data comprises using background n-gram word data and adaptation n-gram word data.

4.   The method of claim 3 wherein generating background n-gram word data comprises generating background n-gram word data for multi-word semantic entities with each data entry comprising a selected number of words.

5.   The method of claim 4 wherein generating adaptation n-gram word data comprises generating adaptation n-gram word data for multi-word semantic entities with each data entry comprising a selected number of words.

6.   The method of claim 4 wherein generating background n-gram class count data based on the background data and the semantic entities and classes thereof comprises tagging word level background data based on the semantic entities and classes thereof.

7.   The method of claim 5 wherein generating adaptation n-gram class count data based on the adaptation data and the semantic entities and classes thereof comprises tagging word level adaptation data based on the semantic entities and classes thereof.

8.   The method of claim 6 wherein generating background n-gram class count data based on the background data and the semantic entities and classes thereof comprises counting unique class level n-grams of the tagged background data.

9.   The method of claim 7 wherein generating adaptation n-gram class count data based on the adaptation data and the semantic entities and classes thereof comprises counting unique class level n-grams of the tagged adaptation data.

10.   The method of claim 8 wherein generating background n-gram class count data based on the background data and the semantic entities and classes thereof comprises discarding some class n-grams from the tagged background data.

11.   The method of claim 9 wherein generating adaptation n-gram class count data based on the adaptation data and the semantic entities and classes thereof comprises discarding some class n-grams from the tagged adaptation data.

12.   A computer-readable medium having computer-executable instructions for performing steps to generate a language model, the steps comprising:

        receiving a set of semantic entities used in a selected domain and organized in classes;

        receiving background n-gram class count data correlated to classes of the set of semantic entities and based on background data indicative of general text;

        receiving adaptation n-gram class count data correlated to classes of the set of semantic entities and based on adaptation data indicative of a selected domain to be modeled; and

        training a language model based on the background n-gram class count data, the adaptation n-gram class count data and the set of semantic entities.

13.   The computer-readable medium of claim 12 wherein training the language model comprises computing background word count data based on the

background n-gram class count data and the set of semantic entities.

14.   The computer-readable medium of claim 13 wherein training the language model comprises computing adaptation word count data based on the adaptation n-gram class count data and the set of semantic entities.

15.   The computer-readable medium of claim 14 wherein training the language model comprises smoothing the n-gram relative frequencies.

16.   The computer-readable medium of claim 15 wherein smoothing comprises using a deleted-interpolation algorithm.